



THE UNIVERSITY *of* EDINBURGH

Edinburgh Research Explorer

## Exploratory Structural Equation Modeling of Personality Data

**Citation for published version:**

Booth, T & Hughes, DJ 2014, 'Exploratory Structural Equation Modeling of Personality Data', *Assessment*, vol. 21, no. 3, pp. 260-271. <https://doi.org/10.1177/1073191114528029>

**Digital Object Identifier (DOI):**

[10.1177/1073191114528029](https://doi.org/10.1177/1073191114528029)

**Link:**

[Link to publication record in Edinburgh Research Explorer](#)

**Document Version:**

Peer reviewed version

**Published In:**

Assessment

**Publisher Rights Statement:**

© Exploratory Structural Equation Modeling of Personality Data. / Booth, Tom; Hughes, David J.  
In: Assessment, Vol. 21, No. 3, 28.03.2014, p. 260-271.

**General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

**Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



Exploratory Structural Equation Modelling of Personality Data

Tom Booth<sup>1</sup> & David John Hughes<sup>2</sup>

<sup>1</sup>Department of Psychology, University of Edinburgh.

<sup>2</sup>Psychometrics at Work Research Group, Manchester Business School, University of  
Manchester.

Correspondence concerning this article should be addressed to Tom Booth, Centre for  
Cognitive Ageing and Cognitive Epidemiology, Department of Psychology, University of  
Edinburgh, 7 George Square, Edinburgh, Scotland, EH8 9JZ.

Email: [tom.booth@ed.ac.uk](mailto:tom.booth@ed.ac.uk)

**FINAL SUBMITTED PDF BEFORE COPY EDITING**

### **Abstract**

The current article compares the use of exploratory structural equation modelling (ESEM) as an alternative to confirmatory factor analytic (CFA) models in personality research. We compare model fit, factor distinctiveness, and criterion associations of factors derived from ESEM and CFA models. In sample 1 (n=336) participants completed the NEO-FFI, the Trait Emotional Intelligence Questionnaire-Short Form and the Creative Domains Questionnaire. In Sample 2 (n=425) participants completed the Big Five Inventory, and the depression and anxiety scales of the General Health Questionnaire. ESEM models provided better fit than CFA models, but ESEM solutions did not uniformly meet cut-off criteria for model fit. Factor scores derived from ESEM and CFA models correlated highly (.91 to .99), suggesting the additional factor loadings within the ESEM model add little in defining latent factor content. Lastly, criterion associations of each personality factor in CFA and ESEM models were near identical in both inventories. We provide an example of how ESEM and CFA might be used together in improving personality assessment.

**Key Words:** ESEM; Confirmatory Factor Analysis; NEO-FFI; BFI; Five Factor Model.

## Introduction

Omnibus personality models, models which build from large item pools, through facets, to higher order factors; have often failed to demonstrate adequate model fit within confirmatory factor analyses (CFA; e.g. Church & Burke, 1994). A number of authors have interpreted these findings as indicative of the inability of CFA to accurately model the complexity of human personality (Borkenau & Ostendorf, 1990; Church & Burke, 1994; McCrae, Zonderman, Costa, Bond, & Paunonen, 1996). In the recent literature, exploratory structural equation modelling (ESEM; Asparouhov & Muthén, 2009) has been suggested as a promising tool in personality research and a possible alternative to CFA (Marsh, et al., 2009; 2010; 2011a; 2011b). The focus of the current study is to compare CFA and ESEM for both structural and criterion analyses of personality data, and to further explore the similarities and differences in the latent constructs which result from the application of each methodology.

CFA and ESEM models differ in two main ways: the use of theory in model specification and the treatment of non-target loadings. Within CFA, all parameters (both included and excluded) are specified by the researcher *a priori* and represent distinct hypotheses about the associations between both observed and latent constructs (Bollen & Pearl, 2012). Often, such models take an independent clusters approach in that each indicator is loaded by only one factor, with all other possible loadings set to zero (see Figure 1, Panel a). In contrast, the only a priori information required to run an ESEM model is the number of factors, all other parameters are freely estimated. That is, in the same manner as an exploratory factor analysis, all factors are allowed to load on all indicators (see Figure 1, Panel b). Interested readers are referred to Asparouhov and Muthén (2009) for an extended technical description of ESEM.

[Insert Figure 1]

CFA is a statistical tool which, theoretically at least, is uniquely suited to model testing (Bollen, 1989; Jöreskog & Sörbom, 1999). CFA also offers a number of additional features which suggest its suitability for the assessment of personality inventories. Firstly, CFA utilizes latent constructs, which retain a theoretical consistency with the prevailing view of personality traits as latent dispositions (Borsboom, 2006; Muliak, 2005). Secondly, CFA models are easily extended to include second order latent constructs, in keeping with the view that personality is hierarchically structured (e.g. the five factor model Costa & McCrae, 1992a; the 16PF Cattell & Cattell, 1995). Thirdly, personality items, generally, share only modest amounts of common variance, thus resulting in higher specific or error variances. CFA explicitly models this specific variance in items and factors as residual variance, resulting in estimates of latent traits comprised of only common or shared variance.

The suitability of CFA models is generally based on an assessment of model fit. Commonly used fit criteria include the root-mean square error of approximation (RMSEA), the standardised root mean square residual (SRMR), comparative fit index (CFI) and Tucker-Lewis or Non-Normed fit index (TLI or NNFI). The values of these indices derived from any particular model are compared against general cut-offs for good fit taken from simulation studies (e.g. Hu and Bentler, 1998; 1999)<sup>1</sup>. However, as previously noted, CFA studies of personality inventories have, more often than not, resulted in models which do not reach conventional criteria for model fit (e.g. studies of the Five Factor Model of Personality: Borkenau & Ostendorf, 1990; Holden & Fekken, 1994; Church & Burke, 1994; Aluja, Garcia, Garcia & Seisdedos, 2005; Hopwood & Donnellan, 2010; Vassend & Skrondal, 2011).

Based on the observation of a lack in model fit, a number of authors have suggested alternative methods for assessing the structural quality of personality psychometrics (e.g. McCrae et al., 1996). Most recently, Marsh et al. (2010) have suggested that ESEM may

provide a more flexible framework for the analysis of the internal structure of personality inventories. The ESEM model identifies exploratory factors (i.e. factors loading on all indicators) within a structural equation modelling framework (for details see Asparouhov & Muthén, 2009). Marsh et al. (2009; 2010; 2011a; 2011b) have argued that the ESEM framework is better suited than CFA to the analysis of complex individual difference constructs, in particular, personality (Marsh et al., 2010) and motivation (Marsh et al., 2011a). Practically, ESEM allows many of the stringent tests of structure which are available within CFA (e.g. measurement invariance), but not exploratory factor analysis, and as a result, it is argued that ESEM allows for rigorous tests of exploratory factors (Asparouhov & Muthén, 2009).

ESEM is argued to address two specific concerns related to CFA analyses of complex individual difference constructs such as personality, namely, the general lack of model fit in CFA studies and the generally large inter-factor correlations in CFA models. With respect to model fit, it is suggested that the fixing of non-target loadings to zero within standard CFA may not be a plausible assumption, and a substantial source of misfit when modelling complex constructs where one may expect to see a large number of cross-loadings (Marsh & Hau, 2007; Marsh et al. 2010), be they substantive or the result of the measurement properties of a given inventory.

In addition, and also as a result of constraining the majority of factor loadings to zero, it is suggested that factors within the CFA model are less distinct, with inter-factor correlations inflated (Marsh et al. 2010). Such a conclusion is based on the results of simulation studies which demonstrate that if factor loadings of moderate magnitude are constrained to be zero, the unmodelled shared association causes an upward bias in the estimate of the factor correlations (Cole, Ciesla & Steiger, 2007). Marsh et al. (2010) argue

that such bias a) leads to less distinct factors and thus blurs conceptual meaning and criterion associations, and b) increases the risk of multicollinearity based estimation problems.

A brief review of the literature provides some support for the proposed advantages of ESEM. In total, a systematic literature search located 42 studies reporting the use of ESEM: 14 of which presented model fit, factor inter-correlations, or both, from comparable ESEM and CFA models (see Table 1). Across all studies which reported factor correlations, both the average, and the single highest correlation were larger for CFA models.

(Insert Table 1 about here)

The picture with regard to model fit is somewhat more mixed. In five studies, both the CFA and ESEM solutions display good fit (Marsh et al. 2011a; Marsh et al., 2011b; Morin & Maiano, 2011; Myers et al., 2011; Sanchez-Carracedo, et al., 2012). In one study (Studerus et al., 2010) both CFA and ESEM showed poor fit. In three studies (Kritjansson et al., 2011; Mattsson, et al., 2012; Meleddu et al., 2012), the CFA solution showed poor fit, whilst the ESEM showed good model fit. In five studies, the CFA models displayed poor fit and adequacy of the ESEM solutions would be dependent on the particular index focussed upon (Furnham et al., 2012; Marsh et al., 2009; 2010; Perry et al., 2012; Rosellini & Brown, 2011). Specifically, the RMSEA indicated fit in both CFA and ESEM models, the CFI indicated a lack of fit for the CFA models and borderline fit for the ESEM models and the TLI evidenced poor fit regardless of the method used. Evidence of poor fit based on the TLI is especially interesting as it suggests that the addition of factor loadings within the ESEM model does not improve the chi-square/degrees of freedom ratio to enough of a degree to suggest fit; or in other words, there is limited gain in fit per additional factor loading included in the model. Interestingly, it is this last group which contains the studies of personality data.

The evidence obtained from this literature review reveals support for the argument that ESEM provides an *improvement* in fit over CFA. However, it is not necessarily the case

that this improvement leads ESEM models to fit according to conventional criteria. In sum, some models fit regardless of the modelling method used, these models are likely well specified models; some models do not fit at all, these are likely poorly specified models; and some models show mixed fit across methods, suggesting further exploratory work as to the structure of the data may be required.

The current study aims to provide further comparisons of ESEM and CFA in the analysis of personality data based on two commonly used five factor model inventories, namely, the NEO-FFI (Costa & McCrae, 1992a) and the Big Five Inventory (BFI: John, Donahue & Kentle, 1991). Our primary focus is on the two substantive concerns with CFA raised by Marsh and colleagues. First, we provide an assessment of model fit based on ESEM and CFA. Next, we explore factor distinctiveness by a) comparing the maximum and mean inter-factor correlations within ESEM and CFA models, b) correlating factor scores for the latent variables derived from both models, and c) comparing criterion associations of the five modelled factors across ESEM and CFA models. Criterion relationships are argued to be important in the assessment of structural models of personality inventories (Hopwood & Donnellan, 2010; Furnham et al., 2012) and are often the primary focus of many studies involving personality. Thus, if the increased correlations observed in CFA do reduce factor differentiation, it is important to understand how, if at all, this affects relationships with external criteria. Information on possible differences in associations will be of benefit to researchers, allowing for informed decisions regarding which methods to apply.

### **Method**

The current study utilised two independent samples, chosen for three main reasons. Firstly, both completed shortened personality inventories and so could be economically modelled at the item level. Secondly, they provided criterion data for which there are well known and differentiated patterns of association, that is, not all of the big five are



hypothesized to be significantly related to the criterion variables. Lastly, the samples come from reasonably distinct populations (adult working population & a student sample) and are of similar size and demographic composition to those commonly used in applied research studies.

## **Sample 1**

### ***Participants***

Sample 1, comprised a total of 425 student participants (307 females & 110 males; 8 participants chose not to disclose their gender) aged between 17 and 65 ( $M = 20.5$ ,  $SD = 5.1$ ). Forty-two percent were European Caucasian; 31.6% were Asian/British Asian; 6.4% African Caribbean; and 15.8% were members of other ethnicities. The majority of participants were single (270, 63.5%), 122 (29.8%) were in a relationship and 32 were married or divorced (7.7%). With regard to educational attainment, 8 participants had been educated to GCSE level, 324 had A-levels, 55 had degrees and 21 had postgraduate degrees (17 participants did not respond).

### ***Measures***

*Personality* was assessed using the 60-item NEO-FFI (Costa & McCrae, 1992a), a well-established, non-timed questionnaire that measures Neuroticism, Extraversion, Openness-to-Experience, Agreeableness, and Conscientiousness. Items involve questions about typical behaviours and are answered on a 5-point Likert-type scale (from strongly disagree to strongly agree). Two-week retest reliability ranges from 0.86 to 0.90 (Robins, Fraley, Roberts, & Trzesniewski, 2001), and internal consistencies range from 0.68 to 0.86 (Costa & McCrae, 1992a).

*Trait Emotional Intelligence (TEI)* was assessed using the TEI Questionnaire–Short Form (TEIQue-SF; Petrides & Furnham, 2006). The 30-item TEIQue-SF which can be scored to produce a total Trait EI score was specifically designed as an efficient measure of global

trait EI. Items were responded to on a 7-point Likert-type scale (from completely disagree to completely agree).

*Creativity* was assessed using the Creative Domains Questionnaire (CDQ; Kaufman, 2006) which measures self-reported Creativity across 56 domains such as Chemistry, Music Composition, and Humour/Comedy. Participants rated their Creativity for each of the domains on a 6-point Likert-type scale ranging from “Not at all Creative” (1) to “Extremely Creative” (6). Participants were also given the opportunity to mark ‘Not Applicable’, which was scored as missing data.

## **Sample 2**

### ***Participants***

The sample consisted of 425 participants. The sample was approximately evenly split by gender (Male = 179, 42%; Female = 246, 58%). Participants reported their age in categories distributed as follows; 18-25 (n=25, 5.9%), 26-34 (n=139, 32.7%), 35-44 (n=116, 27.3%), 45-54 (n=85, 20.0%), 55-64 (n=57, 13.4%) and 65+ (n=3, 0.7%). The majority of the participants were aged between 26 and 44 (n=255, 60.0%). All participants were currently in employment in a variety of organisations in the public and private sector. As data were collected from organisational settings, data on the ethnic background of participants was not collected.

### ***Measures***

*Personality* was assessed using the Big Five Inventory (BFI, John, Donahue & Kentle, 1991). The BFI is a 44-item measure of the five factor model (FFM), measuring Neuroticism (eight items), Extraversion (eight items), Openness (ten items), Agreeableness (nine items) and Conscientiousness (nine items) at the domain level. Participants responded on a five point scale ranging from *Strongly Disagree* to *Strongly Agree*. John and Srivastava (1999) report an average scale Cronbach’s alpha of .83.

*Depression* and *Anxiety* were measured using the 28-item General Health Questionnaire (GHQ-28, Goldberg & Hillier, 1979; Werneke, Goldberg, Yalcin & Uston, 2000). Participants responded on a four point scale (1 = not at all, 2 = no more than usual, 3 = rather more than usual, 4 = much more than usual), equivalent to that used in the full GHQ. The GHQ-28 has scale reliabilities ranging from 0.71 to 0.85 (Vallejo, Jordán, Díaz, Comeche & Ortega, 2007).

### **Analysis Strategy**

All models in the current analysis were run using robust maximum likelihood (MLR) in *Mplus* 6.0. Given that the current data are drawn from inventories with likert-type response formats, and are thus ordinal not continuous, use of weighted least squares mean and variance (WLSMV) adjusted estimation would have been optimal. However, in the current study we wished to compare non-nested models based on AIC and BIC which are not calculable from WLSMV estimation. Accordingly, we selected MLR. For information, we re-ran all models in the current analysis using WLSMV and present the results in supplementary material A. The analysis comprised two major stages.

In Stage 1, we considered the issue of model fit. Item level CFA and ESEM models for both the BFI and NEO-FFI were examined. CFA models were specified in accordance with the original measure publications. Target loadings were freely estimated and non-target loadings were fixed to zero, as is the convention in CFA models. The variances of the latent variables were fixed to unity to identify each model (Bollen, 1989). ESEM models were estimated by specifying five factors for both inventories, using Geomin oblique rotations as recommended by Asparouhov and Muthén (2009).

Model fit was assessed by applying the commonly applied cut-off values of  $>.95$  for the CFI and TLI (Hu & Bentler, 1998; 1999),  $<.06$  for the RMSEA (Browne & Cudeck, 1993),  $<.05$  for the SRMR (Spence, 1997) and an approximate guide of  $<1.0$  for the weighted

root mean residual (WRMR). For the MLR analyses we also report AIC and BIC. For both AIC and BIC, smaller values indicate better models. Raftery (1995) has suggested that differences in BIC of greater than 10 are indicative practically significant improvements in model fit. We use these general criteria here.

In Stage 2, we considered factor distinctiveness. Firstly we compared the inter-factor correlation matrix, mean, and single highest factor correlations in both solutions. Secondly, factor scores were estimated for factors derived using both ESEM and CFA. These factors were then correlated in order to assess the degree of relatedness and proportion of shared variance. The magnitude of the correlations between the CFA and ESEM factors provided an approximate metric of the extent to which the additional loadings in the ESEM model contribute to the definition of the latent construct.

Lastly, in order to examine the differences in personality-criterion relationships across methods, structural models were estimated for each inventory using CFA and ESEM. In each model, the criterion variables were regressed onto the latent personality factors. The magnitude of the regression coefficients was evaluated in order to ascertain if the additional factor loadings specified in ESEM influenced the criterion relationships.

## **Results**

As noted above, though we present the results from the MLR analyses here, all models were also estimated based on WLSMV (see Supplementary Material A). Although there were some small differences in the parameter estimates across estimation methods, the substantive conclusions were consistent.

### ***Model Fit***

Table 2 contains the model fit for the NEO-FFI and BFI based on item level CFA and ESEM models. In both models, the ESEM solution provided better model fit than the CFA model. However, in both cases, the ESEM model failed to provide good fit in accordance

with conventional ranges of suggested cut-off values. The results are consistent with those observed in previous research discussed above and summarised in Table 1.

(Insert Table 2 about here)

The pattern of results for the AIC and BIC are of particular interest. The AIC suggests that the ESEM model is preferable for both the NEO-FFI and BFI. However, the reverse was true for the BIC, according to which the CFA model is preferable, and in the case of the NEO-FFI, this difference was greater than 10, which has been suggested to indicate substantive improvement.

### ***Factor Distinctiveness***

In order to investigate the level of factor distinctiveness across models, we considered the factor inter-correlations, mean, and absolute maximum correlations in the NEO-FFI and BFI (Table 3 and 4 respectively), and the correlations between the factor scores of corresponding personality factors (e.g. CFA Neuroticism with ESEM Neuroticism) from the same inventory derived across methods (Table 5). Again, in keeping with past findings and as anticipated, the single greatest correlation and overall mean correlations were larger in CFA than ESEM for both the NEO-FFI (CFA absolute mean  $r = .25$ ; ESEM absolute mean  $r = .11$ ) and BFI (CFA absolute mean  $r = .25$ ; ESEM absolute mean  $r = .14$ ). However, the highest correlation of -0.47 is not considered to be of the magnitude to suggest excessive multicollinearity (Grewal, Cote & Baumgartner, 2004), although values of this magnitude have been argued to result in suppression effects when predicting certain criterion variables (Beckstead, 2012).

(Insert Table 3 about here)

(Insert Table 4 about here)

(Insert Table 5 about here)

As shown in Table 5, the correlations between factor scores derived from ESEM and CFA ranged from 0.91 to 0.99 for the NEO-FFI, and from 0.92 to 0.99 for the BFI. The factor score determinacies for the ESEM and CFA models for both the BFI and NEO-FFI were high and differed in most cases only at the third decimal place. As such, the estimated scores produced by the two measurement models appear to be equivalently reliable. NEO-FFI Agreeableness showed the largest difference in determinacy estimates (.024). These results suggest the empirical content of scores derived from the two estimation methods are largely identical for Neuroticism, Extraversion, Openness and Conscientiousness, but that there is a small amount of variability in the definition of Agreeableness across estimation methods and inventories. However, correcting the correlations between factor scores derived from ESEM and CFA for the score determinacies (reliabilities) resulted in correlations of unity in all cases. Thus, even in the case of Agreeableness, discrepancies in factor content were small.

To further investigate the distinctiveness of the factors derived across the two methods we considered the criterion associations with CFA and ESEM derived latent factors. Tables 6 (NEO-FFI) and 7 (BFI) contain the standardized parameter estimates and 95% confidence intervals for structural models based on ESEM and CFA modelling of personality factors.

(Insert Table 6 about here)

(Insert Table 7 about here)

The variance explained in the criterion variables is very similar across models. All criterion associations fall within the confidence intervals for the equivalent parameters in the alternative model. Taken collectively, this suggests that ESEM factors are no more distinct than CFA factors when criterion associations are used as a proximal measure of distinctiveness.<sup>2</sup>

### **Exploring the Factor Content of Agreeableness in the NEO-FFI**

As noted above, the NEO-FFI Agreeableness factor showed the weakest correlations across modelling methods ( $r = .91$ , Table 5;  $r = .87$  Supplementary material A, T5). In addition, the regression paths in the structural models for Creativity were significant for CFA but not for ESEM. Thus, the methodology used to estimate the measurement model would in this instance affect the substantive conclusions drawn regarding criterion associations.

To explore this further, all factor loadings which reached significance at  $p < 0.001$  in the ESEM model were added to the CFA solution. We then reassessed model fit, factor distinctiveness and criterion associations.

In total, four items had significant ( $p < 0.05$ ) non-target cross loadings on Agreeableness in the NEO-FFI ESEM model that were consistent across the MLR and WLSMV estimation. Two items had their primary loadings on Extraversion (Item 12: I don't consider myself especially "light-hearted"; Item 27: I usually prefer to do things alone), one on Openness (Item 33: I seldom notice the moods or feelings that different environments produce) and one on Neuroticism (Item 36: I often get angry at the way people treat me). These cross loadings appear to largely represent the overlap in content between Extraversion and Agreeableness (see Discussion).

Inclusion of these four additional loadings in the CFA model led to a very small improvement in model fit ( $\chi^2 = 3592.999$ ; CFI = .686; TLI = .673; RMSEA = .055; SRMR = .083;  $\Delta\chi^2 = 74.038(4)$ ,  $p < .001$ ;  $\Delta$ CFI = .011;  $\Delta$ TLI = .012;  $\Delta$ RMSEA = -.001;  $\Delta$ SRMR = -.002), a small decrease in both maximum ( $r = -.45$ ,  $\Delta.02$ ) and average ( $r = .23$ ,  $\Delta.02$ ) inter-factor correlation, and an increase in the correlations between ESEM and CFA derived factor scores ( $r = .947$ ; 89.68% shared variance). In addition, the parameter estimates showing the relation between Agreeableness in the revised CFA model and both TEI and Creativity increased (-.07 and .21 respectively). In the case of Creativity, the coefficient from the revised CFA was also significant ( $p < .01$ ). Thus in this instance, any substantive interpretation based on purely

reaching a p-value would be equivalent across ESEM and CFA. Therefore, the inclusion of just 4 additional factor loadings in the CFA model produced an Agreeableness factor which was highly consistent with that derived from ESEM.

### **Discussion**

In the current paper we sought to clarify understanding of the differences between models of personality derived from ESEM and CFA. ESEM and CFA solutions were examined in relation to three important issues, namely, model fit, factor content and distinctiveness, and criterion associations. The results revealed that when compared with CFA models, ESEM solutions produce improved overall model fit, reduced inter-factor correlations and similar criterion associations, suggesting that ESEM solutions do have value in the study of personality. However, our results also reveal a number of nuances that are important to consider, especially if making a choice between ESEM and CFA.

#### **Model Fit in CFA and ESEM**

It has been suggested that the misfit observed in CFA models of personality is due, in the main, to the overly restrictive constraint that non-target loadings should be equal to zero. In which case, ESEM models which allow all loadings to be freely estimated provide an interesting alternative and ESEM models should evidence fit where CFA models do not. When examining extant literature and the results of the current study, it is clear that ESEM consistently provides an improvement in model fit over CFA. Importantly however, the increase does not provide categorical evidence of model fit in all circumstances. When ESEM models demonstrate adequate levels of fit, more often than not, so do CFA models. In general, the evidence of misfit from CFA is accompanied by misfit in ESEM. Thus, the levels of fit observed appear to be driven not by the method of analysis applied, but by the quality and accuracy of model specification (e.g. number of factors). In other words, good models fit regardless of whether ESEM or CFA is used.



Further, in the current study we considered additional fit indices to previous studies of ESEM and CFA, namely the AIC and BIC. These indices are especially suited to the comparison of non-nested models and both include penalties for model parsimony, a key issue in the CFA-ESEM discussion. Results were equivocal. The AIC showed preference for the ESEM models, whilst the BIC showed preference for the CFA models. However, as the AIC and BIC have not been regularly reported in prior comparative studies of CFA and ESEM, it is not possible to assess how far these findings generalise.

### **Factor Correlation and Criterion Association in CFA and ESEM**

With regard to factor content and distinctiveness, the current results reveal that the conclusions reached will depend upon the criterion used to judge similarity. The findings in the current study, and the review of the extant literature, support the conclusion that both the average and maximal inter-factor correlations are lower in ESEM than in CFA models. Taking this metric alone, it would be concluded that ESEM factors are more distinct than CFA factors. However, despite the reduction in inter-factor correlations, it must be noted that the factors themselves are shown to be highly similar (average equivalent  $r = 0.96$ , shared variance = 92%) suggesting that the two solutions provide factors that are almost equivalent and thus by definition cannot vary too greatly in terms of distinctiveness.

Nevertheless, the evidence of higher correlations obtained from CFA models could suggest that CFAs are likely to result in increased problems with multicollinearity. Multicollinearity is a problem in regression, and thus, SEM analyses as it can result in the distortion of regression coefficients (Beckstead, 2012). The assertion that CFA may produce problematic levels of multicollinearity was not supported by the current study. Rather, the results reveal minimal differences in the magnitude of variance explained and the profile of standardised beta estimates between CFA and ESEM criterion associations. In the one instance where a cross method discrepancy was observed, the addition of just four additional

parameters produced closely aligned results. However, it is important to note that we did not conduct any direct tests of the effect of multicollinearity in the current study (see for example Beckstead, 2012).

### **An Outstanding Issue; Inter-factor correlations in CFA and ESEM**

At one level, the evidence suggesting the similarity in factor content and criterion associations from the current study may lead one to conclude that either ESEM or CFA may be applied to personality data without further consideration. However, the models still differ, quite substantially, with respect to the inter-factor correlations. Inter-factor correlations are, in and of themselves, of great significance to current practice in personality trait psychology. Personality factor models are, generally speaking, hypothesized to be hierarchical, with higher order constructs (broad factors) hypothesized to be the common latent cause of lower order constructs (facets), which are themselves the common cause of item responses. Debates are on-going as to the number of factors at each level of this hierarchy, whether this be between the Eysenkian gigantic 3 or the big five (e.g. Costa & McCrae, 1992b; Eysenck, 1991, 1992); or whether it be the existence (or not) of the Big Two or general factor of personality (e.g. DeYoung, 2006; Irwing, 2013; Reimann & Knadler, 2010); or the existence of intermediate levels of hierarchy (e.g. DeYoung, Quilty & Peterson, 2007). All such structural debates are reliant on the accurate measurement of inter-item and inter-factor correlations. How then should we interpret and think about the observation of higher correlations between factors in CFA versus ESEM models?

To date, the observation that correlations are larger in CFA than ESEM has been taken as evidence that the CFA estimates are biased (Marsh et al., 2010), such that when factor loadings of moderate magnitude are constrained to zero, there is an upward bias on the inter-factor correlations. Within SEM, factor loadings and inter-factor correlations offer two different modelling options to represent the variance of observed indicators. For a factor

loading to be preferred over a slightly larger inter-factor correlation which may result from constraining this parameter to zero, it must be demonstrable that the factor loading is substantive.

Current evidence suggests that it is questionable whether all such cross-loadings within ESEM and EFA models of personality can be argued to be substantive. We suggest this on a number of grounds. Specifically, we consider a substantive loading to be one of reasonable magnitude, which directly contributes to the definition of the latent construct in a theoretically consistent way. The results of the current study point to the fact that the additional factor loadings do little to define the constructs of interest and are small in magnitude. Thus, it is difficult to argue that such loadings are substantive.

### **Do cross-loadings capture complexity?**

One possibility in arguing for the modelling of all possible factor loadings is to suggest that these are required in order to capture the complexity of personality. We believe there is an important distinction in this discussion, namely, that between measurement and theoretical complexity. Personality is undoubtedly theoretically complex, but this does not mean that the measurement complexity observed in the application of latent models is due to this theoretical complexity. Measurement complexity can, and likely is, also caused by a wealth of questionnaire design issues such as poor and ambiguous item wording, overly complex items, and a lack of understanding of the relation between rating scales and items.

To the extent that the lack of fit in CFA is a marker of measurement complexity, then ESEM is a highly useful tool, perhaps more so than standard EFA approaches due to the modelling of residual error, and the possibility to conduct advanced analyses such as investigating measurement invariance (Marsh et al., 2010). As has been suggested by others and is repeated here, ESEM may be especially useful during the development of psychometric instruments. However, we suggest that ESEM should be used with a view to, as

far as possible, identifying and removing atheoretical complexity whilst retaining that which is meaningful, with CFA remaining a reasonable *gold standard* aim for measurement. For example, if items are hypothesized to be complex and to measure multiple aspects of human personality, such paths can be specified a priori, and tested for their plausibility and consistency within a CFA framework. Our application of ESEM and CFA to the analysis of Agreeableness may, to some small degree, show the utility of combining methods to this end.

### **CFA and ESEM working in harmony for questionnaire development**

The analysis of the Agreeableness factor of the NEO-FFI provides an example of how ESEM and CFA may be used alongside one another in the analysis of measurement complexity. The Agreeableness factor of the NEO-FFI showed the weakest correlation across the two methods (CFA versus ESEM). In the criterion models, Agreeableness was the only factor for which the beta coefficients across methods showed any marked differences. Therefore, by the criteria used in the current paper, ESEM and CFA provided Agreeableness factors which could be argued to have substantively different interpretations. However, the inclusion of just four additional factor loadings, primarily from items hypothesized to load on Extraversion, led to largely comparable Agreeableness factors from ESEM and CFA in terms of the factor score and beta coefficients. This questions exactly what the additional 42 parameters included within the ESEM model contribute.

The additional Extraversion items consider the extent to which someone is light-hearted and cheerful. Further, the additional Neuroticism loading on Agreeableness concerns the expression of anger. Such item content appears theoretically justifiable based on a broad definition of Agreeableness as altruistic, gentle, kind, cooperative, straightforward, warm, and sympathetic (Costa & McCrae, 1992a). Therefore, the inclusion of a number of cross loadings based on this content overlap could be specified a priori within a CFA model. The inclusion of such cross loadings would violate the independent clusters assumption noted to

be problematic in CFA studies (Marsh et al., 2010), but would not, for example, violate the definition of simple structure as originally proposed by Thurstone (1947). Further, the inclusion of a small number of a priori specified cross loadings is in the mind of the current authors preferable to the inclusion of a large number of small, non-significant, and atheoretical factor loadings.

### **Some concluding thoughts**

To some extent, whether one agrees or disagrees with ESEM as a modelling strategy is dependent on the importance placed on different forms of model misspecification. Hu and Bentler (1998) discuss two types of misspecification in structural models: omitting substantively meaningful parameters, and the inclusion of substantively irrelevant parameters. Saris, Satorra and van der Veld (2009) argue that it is the former that is most problematic within structural modelling. From this perspective, one would prefer to use ESEM as the inclusion of all measurement parameters ensures that no substantive parameters are omitted. However, the inclusion of all possible parameters could be interpreted as overly inclusive and certainly contradictory to the principle of parsimony. Further, the results of the current study suggest that in the majority of instances, the additional parameters add little to the definition of the latent constructs as evidenced by both the correlations between factor scores derived from ESEM and CFA, and the criterion relationships. Taken collectively, these findings pose questions regarding the substantive and theoretical importance of these parameters.

Marsh and colleagues have stated that “*the traditional [independent clusters CFA] model is not appropriate for the NEO-FFI and suspect that this would also be the case for many personality measures*”. We agree. It is clear that, as currently specified, personality models do not fit an independent clusters model. Marsh and colleagues argue further that their results showing improved fit in ESEM and smaller inter-factor correlations “*provide*

*clear evidence that an ESEM approach is more appropriate than a traditional [independent clusters CFA] approach for Big Five responses”* (Marsh et al., 2010 p. 489). In this instance, we do not fully agree. On the basis of the literature review conducted, the current results, and consideration of personality research in general, whether ESEM is more appropriate than CFA is debatable and a number of additional considerations are necessary.

First, whilst, compared to CFA, ESEM models of personality provide better fit; they still do not fit according to conventional criteria. This picture is further complicated when model fit indices with parsimony corrections are considered. So, if one rejects CFA on the basis of fit, in many cases one must also reject ESEM.

Second, in terms of factor distinctiveness, the picture is not at all clear. Inter-factor correlations are discrepant across modelling methods whilst factor content and criterion associations are largely consistent. Thus, we follow Hopwood and Donnellan (2010) in suggesting that the abandonment of CFA as a methodology for the assessment of personality structure is too extreme.

Third, ESEM and CFA are tools designed for different purposes; the former, like EFA, for exploration and the latter for confirmation or hypothesis testing. Thus, the choice between the two should, at least in part, be based on specific research goals. To the extent that measures of personality represent ‘work in progress’, ESEM is warranted, but if personality models are considered ‘complete’ then they should be tested using CFA. Whether that model is based on an independent clusters approach or not is a question for the specific theory being tested. If the CFA based evidence fails to support the theoretically espoused model, then ESEM could be invaluable during a period of iterative model development.

In sum, the current study has examined the utility of ESEM and CFA as tools for the modelling of personality data. The study has shown that in most cases where ESEM models achieve adequate fit, so do equivalent CFA models. In instances where ESEM models fail to

fit, so do CFA models. Thus, the often observed misfit in personality models does not appear to be driven by the tool of analysis but instead by model misspecification, misspecification that goes beyond constraints placed on non-target loadings. Further, it has been shown that ESEM and CFA derived factors are less different than previously suggested. In the main, equivalent factors derived from ESEM and CFA share a large proportion of variance and display highly similar relationships with criterion variables. With the exception of an increment in model fit, there appears to be little clear evidence supporting the application of ESEM over CFA. This is especially true when one considers the theoretical implications of modelling all possible parameters and the different goals of the two methods. Thus, whilst ESEM can be a useful tool during model building and development, suggestions that ESEM should be used in place of CFA appear premature if not unwarranted. Rather, we see the most fruitful approach being to use both methods in order to improve personality assessment.

### References

- Aluja, A., Garcia, O., Garcia, L.F. & Seisdedos, N. (2005). Invariance of the “NEO-PI-R” factor structure across exploratory and confirmatory factor analyses. *Personality and Individual Differences*, 38, 1879-1889.
- Asparouhov, T. & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling*, 16, 397-438.
- Barrett, P. (2007), Structural Equation Modelling: Adjudging Model Fit. *Personality and Individual Differences*, 42, 815-24.
- Beckstead, J. W. (2012). Isolating and examining sources of suppression and multicollinearity in multiple linear regression. *Multivariate Behavioral Research*, 47, 224-246.
- Bollen, K.A. (1989). *Structural Equations with Latent Variables*. New York: Wiley.

- Bollen, K.A. & Pearl, J. (2012). Eight Myths about Causality and Structural Equation Models. In S. Morgan (Eds.) *Handbook of Causal Analysis for Social Research*. Springer.
- Borkenau, P., & Ostendorf, F. (1990). Comparing exploratory and confirmatory factor analysis: A study on the 5-factor model of personality. *Personality and Individual Differences*, 11, 515-524.
- Borsboom, D. (2006). The attack of the psychometricians. *Psychometrika*, 71, 425-440.
- Browne, M.W., & Cudeck, R. (1993). Alternative Ways of Assessing Model Fit. In K. A. Bollen and J. S. Long (Eds). *Testing structural equation models*. pp.136-162. London, UK. Sage Publications.
- Cattell, R.B., & Cattell, H.E.P. (1995). Personality Structure and the new Fifth Edition of the 16PF. *Educational and Psychological Measurement*, 55, 926-937.
- Church, T.A., & Burke, P.J. (1994). Exploratory and Confirmatory Tests of the Big Five and Tellegen's Three- and Four-Dimensional Models. *Journal of Personality and Social Psychology*, 66, 93-114.
- Cole, D.A., Ciesla, J.A., & Steiger, J.H. (2007). The insidious effects of failing to include design-driven correlated residuals in latent-variable covariance structure analysis. *Psychological Methods*, 12, 381-398.
- Costa, P.T. & McCrae, R.R. (1992a). *Revised NEO Personality Inventory (NEO-PI-R) and NEO Five-Factor Inventory (NEO-FFI) professional manual*. Odessa, FL: Psychological Assessment Resources.
- Costa, P. T., & McCrae, R. R. (1992b). Four ways five factors are basic. *Personality and individual differences*, 13, 653-665.
- Cramer, A.O.J., van der Sluis, S., Noordhof, A., Wichers, M., Geschwind, N., Aggen, S.H., Kendler, K.S., & Borsboom, D. (2012). Dimensions of Normal Personality as



- Networks in Search of Equilibrium: You Can't Like Parties if You Don't Like People. *European Journal of Personality*, 26, 414-431.
- Cramer, A.O.J., Waldrop, L.J., van der Mass, H.L.J., & Borsboom, D. (2010). Complex realities require complex theories: Refining and extending the network approach to mental disorders. *Behavioral and Brain Sciences*, 33, 178-193.
- DeYoung, C. G. (2006). Higher-order factors of the Big Five in a multi-informant sample. *Journal of personality and social psychology*, 91, 1138-1151.
- DeYoung, C. G., Quilty, L. C., & Peterson, J. B. (2007). Between facets and domains: 10 aspects of the Big Five. *Journal of personality and social psychology*, 93 880-896.
- Eysenck, H. J. (1991). Dimensions of personality: 16, 5 or 3?—Criteria for a taxonomic paradigm. *Personality and individual differences*, 12, 773-790.
- Eysenck, H. J. (1992). Four ways five factors are not basic. *Personality and individual differences*, 13, 667-673.
- Furnham, A., Guenole, N., Levine, S., & Chamorro-Premuzic, T. (2012). The NEO-PI-R: Factor structure and gender invariance from exploratory structural equation modelling analyses in a high-stakes setting. *Assessment*.
- Goldberg, D. P., & Hillier, V. F. (1979). A scaled version of the General Health Questionnaire. *Psychological medicine*, 9, 139-145.
- Grewal, R., Cote, J.A., & Baumgartner, H. (2004). Multicollinearity and Measurement Error in Structural Equation Models: Implications for Theory Testing. *Marketing Science*, 23, 519-529.
- Holden, R.R., & Fekken, G.C. (1994). The NEO five-factor inventory in a Canadian context: Psychometric properties for a sample of university women. *Personality and Individual Differences*, 17, 441-444.

- Hopwood, C.J., & Donnellan, M.B. (2010). How should the internal structure of personality inventories be evaluated? *Personality and Social Psychology Review*, 14, 332-346.
- Hu, L.T., & Bentler, P.M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods*, 3, 424-453.
- Hu, L.T., & Bentler, P. M. (1999). Cut-off criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1-55.
- Irwing, P. (2013). A general factor of personality: Substance or artefact? *Personality and Individual Differences*, 55, 234-242.
- John, O. P., Donahue, E. M., & Kentle, R. L. (1991). *The Big Five Inventory-Versions 4a and 54*. Berkeley, CA: University of California, Berkeley, Institute of Personality and Social Research.
- John, O.P. & Srivastava, S. (1999). The Big-Five Trait Taxonomy: History, Measurement, and Theoretical Perspectives. In L. Pervin & O.P. John (Eds.), *Handbook of Personality: Theory and Research (2nd ed.)*. New York: Guildford.
- Jöreskog, K. G., & Sörbom, D. (1999). LISREL 8.30. *Chicago: Scientific Software International*.
- Kaufman, J. C. (2006). Self-reported differences in creativity by gender and ethnicity. *Journal of Applied Cognitive Psychology*, 20, 1065-1082.
- Krisjansson, S. D., Pergadia, M. L., Agrawal, A., Lesov-Schlaggar, C. N., McCarthy, D. M., Piasecki, T. M., Duncan, A. E., Bucholz, K. K., Madden, P. A., Sher, K. J., & Heath, A. C. (2011). Smoking outcome expectancies in young adult female smokers: Individual differences and associations with nicotine dependence in a genetically informative sample, *Drug and Alcohol Dependence*, 116, 37-44

- Marsh, H. W., & Hau, K-T. (2007). Applications of latent-variable models in educational psychology: The need for methodological-substantive synergies. *Contemporary Educational Psychology*, 32, 151-171.
- Marsh, H. W., Lüdtke, O., Muthén, B., Asparouhov, T., Morin, A. J. S., Trautwein, U. & Nagengast, B. (2010). A new look at the big-five factor structure through exploratory structural equation modeling. *Psychological Assessment*, 22, 471-491.
- Marsh, H.W., Muthén, B., Asparouhov, T., Ludtke, O., Robitzsch, A., Morin, A.J.S., & Trautwein, U. (2009). Exploratory structural equation modeling, integrating CFA and EFA: Application to students' evaluations of university teaching. *Structural Equation Modeling*, 16, 439-476.
- Marsh, H.W., Liem, G.A.D., Martin, A.J., Morin, A.J.S., & Nagengast, B. (2011). Methodological Measurement Fruitfulness of Exploratory Structural Equation Modeling (ESEM): New Approaches to Key Substantive Issues in Motivation and Engagement, *Journal of Psychoeducational Assessment*, 29, 332-346.
- Marsh, H.W., Nagengast, B., Morin, A.J.S., Parada, R.H., Craven, R.G., & Hamilton, L.R. (2011). Construct Validity of the Multidimensional Structure of Bullying and Victimization: An Application of Exploratory Structural Equation Modeling, *Journal of Educational Psychology*, 103, 701-732
- Mattsson, M. (2012). Investigating the factorial invariance of the 28-item DBQ across genders and age groups: An Exploratory Structural Equation Modeling Study, *Accident Analysis and Prevention*, 48, 379-396.
- McCrae, R.R., Zonderman, A.B., Costa, P.T., Bond, M.H. & Paunonen, S.V. (1996). Evaluating Replicability of Factors in the Revised NEO Personality Inventory: Confirmatory Factor Analysis Versus Procrustes Rotation. *Journal of Personality and Social Psychology*, 70, 552-566.

- Meleddu, M., Guicciardi, M., Scalas, F., & Fadda, D. (2012). Validation of an Italian version of the Oxford Happiness Inventory in adolescence, *Journal of Personality Assessment*, 94, 175-185.
- Morin, A.J.S., & Maiano, C. (2011). Cross-validation of the short form of the physical self-inventory (PSI-S) using exploratory structural equation modeling (ESEM), *Psychology of Sport and Exercise*, 12, 540-554
- Mulaik, S. (2005). Looking back on the indeterminacy controversies in factor analysis. In A. Maydeu-Olivares, and J. J. McCardle (eds.), *Contemporary psychometrics*. Mahwah: NJ: Lawrence Erlbaum.
- Myers, N.D., Chase, M.A., Pierce, S.W., & Martin, E. (2011). Coaching efficacy and exploratory structural equation modelling: A substantive-methodological synergy. *Journal of Sport & Exercise Psychology*, 33: 770–806.
- Perry, J.L., Clough, P.J., Crust, L., Earle, K., & Nicholls, A.R. (2013). Factorial validity of the Mental Toughness Questionnaire-48, *Personality and Individual Differences*, 54, 587-592.
- Petrides, K. V. & Furnham, A. (2006). The role of trait emotional intelligence in a gender-specific model of organizational variables. *Journal of Applied Social Psychology*, 36, 552-569.
- Raftery, A. E. (1995). Bayesian model selection in social research. *Sociological Methodology*, 25, 111-164.
- Riemann, R., & Kandler, C. (2010). Construct validation using multitrait - multimethod - twin data: The case of a general factor of personality. *European Journal of Personality*, 24, 258-277.
- Robins, R. W., Fraley, R. C., Roberts, B. W., & Trzesniewski, K. H. (2001). A longitudinal study of personality change in young adulthood. *Journal of Personality*, 69, 617–640.

- Rosellini, A.J., & Brown, T.A. (2011). The NEO Five-Factor Inventory: Latent Structure and Relationships With Dimensions of Anxiety and Depressive Disorders in a Large Clinical Sample, *Assessment*, 18, 27-38
- Sánchez-Carracedo, D., Barrada, J. R., López-Guimerà, G., Fauquet, J., Almenara, C. A., & Trepát, E. (2012). Analysis of the factor structure of the Sociocultural Attitudes Towards Appearance Questionnaire (SATAQ-3) in Spanish secondary students through exploratory structural equation modelling. *Body Image*, 9, 163–171.
- Saris, W.E., Satorra, A., & van der Veld, W.M. (2009). Testing structural equation models or detection of misspecifications? *Structural Equation Modeling*, 16, 561-582.
- Spence, S.H. (1997). Structure of Anxiety Symptoms Among Children: A confirmatory factor-analytic study. *Journal of Abnormal Psychology*, 106, 280-297.
- Studerus E, Gamma A, Vollenweider FX (2010): Psychometric evaluation of the altered states of consciousness rating scale (OAV). *PLoS One* 5:e12412
- Vallejo, M.A., Jordán, C.M., Díaz, M.I., Comeche, M.I., y Ortega, J. (2007). Psychological Assessment via the Internet: A reliability and validity study of online (vs Paper-and-Pencil) Versions of the General Health Questionnaire-28 (GHQ-28) and the Symptoms Check-List-90-Revised (SCL-90-R). *Journal of Medical Internet Research*, 9(1): e2.
- Vassend, O., & Skrondal, A. (2011). The NEO personality inventory revised (NEO-PI-R): Exploring the measurement structure and variants of the five-factor model. *Personality and Individual Differences*, 50, 1300-1304.
- Werneke, U., Goldberg, D. P., Yalcin, I., & Ustun, B. T. (2000). The stability of the factor structure of the General Health Questionnaire. *Psychol Med*, 30, 823-829.

Footnotes

<sup>1</sup> Although it is common for confirmatory factor models to be assessed based on model fit indices and cut-off values, it remains a much debated topic. For example, see the target article by Barrett (2007) and the associated commentaries.

<sup>2</sup> In the analyses based on WLSMV estimation presented in supplementary material A, the regression coefficients from Agreeableness to both TEI and Creativity differed significantly across models. The correlations between the factor scores for Agreeableness across the CFA and ESEM solutions were also lower ( $BFI = .88$ ;  $NEO-FFI = .87$ ). Thus the differences between the CFA and ESEM Agreeableness factors were slightly larger under WLSMV estimation.

Table 1

Comparative CFA and ESEM Results from Published Research Studies

Study	CFA Fit			Correlation		ESEM Fit			Correlation	
	RMSEA	CFI	TLI	Mean	Max	RMSEA	CFI	TLI	Mean	Max
<i>Personality</i>										
Marsh et al. (2010)*	0.044	0.750	0.731	.197	-.500	0.028	0.914	0.893	.064	-.205
Rosellini & Brown (2011)**	0.05/0.05	0.80/0.79	0.78/0.77	N/A	N/A	0.05/0.04	0.91/0.90	0.89/0.88	.32/.27	.12/.16
Furnham, Guenole, Levine & Chamorro-Premuzic (2012)	0.11	0.65	0.60	N/A	N/A	0.06	0.91	0.86	N/A	N/A
<i>Other</i>										
Kristjansson et al. (2011)	0.174	0.883	0.959	N/A	N/A	0.031	0.994	0.997	.38	.61
Marsh et al. (2011a)	0.033	0.935	0.928	.40	.77	0.025	0.977	0.958	.17	.38
Marsh et al. (2011b)	0.029	0.943	0.938	.44	.84	0.025	0.963	0.955	.29	.53
Marsh, Muthén, Asparouhov, Lüdtke, Robitzsch, Morin & Trautwein (2009)	0.111	0.887	0.871	.59	.86	0.084	0.961	0.927	.33	.53
Mattsson (2012)	0.069	0.861	0.929	.57	.80	0.036	0.960	0.981	.36	.61
Meleddu, Guicciardi, Scalas & Fadda (2012)***	0.064	0.883	0.858	N/A	N/A	0.039	0.931	0.912	N/A	N/A
Morin & Maiano (2011)	0.053	0.948	0.934	.65	.92	0.034	0.990	0.973	.33	.51
Myers, Chase, Pierce & Martin (2011)	0.075	0.969	0.972	.75	.88	0.033	1.00	0.99	.48	.71
Sanchez-Carracedo, Barrada, Lopez-Guimera, Fauquet, Almenara, Trepas (2012)	0.055	0.931	0.917	N/A	N/A	0.047	0.960	0.940	.387	.637
Perry, Clough, Crust, Earle & Nicholls (2013)	0.046	0.853	0.845	.78	.90	0.033	0.942	0.924	N/A	N/A
Studerus, Gamma & Vollenweider (2010)	0.061	0.738	0.729	N/A	N/A	0.050	0.777	0.754	N/A	N/A

---

\* The authors report only median correlations.

\*\* The authors report on two samples. The models reported include correlated residuals between items from the same NEO facet scales.

\*\*\* Results reported here are taken from the 3 factor solution reported by the authors.



Table 2: Model Fit for Five Factor Models of the NEO-FFI (n=336) and BFI (n=425)

	$\chi^2$	<i>df</i>	Sig.	CFI	TLI	RMSEA	WRMR/ SRMR	AIC	BIC
<i>NEO-FFI</i>									
5 Factor CFA	3667.037	1700	<i>p</i> <.001	0.675	0.661	0.056	0.085	57176.397	57917.378
5 Factor ESEM	2658.034	1480	<i>p</i> <.001	0.805	0.767	0.047	0.044	56477.169	58076.127
<i>BFI</i>									
5 Factor CFA	2347.872	892	<i>p</i> <.001	0.739	0.723	0.062	0.080	51005.036	51580.433
5 Factor ESEM	1498.985	736	<i>p</i> <.001	0.863	0.824	0.049	0.040	50379.748	51587.270

Table 3: Factor Inter-correlations for the NEO-FFI Model from CFA and ESEM Models

	1	2	3	4	5
1. Neuroticism	-	-.211 <sup>***</sup>	-.102	-.028	-.182 <sup>**</sup>
2. Extraversion	-.469 <sup>***</sup>	-	.098	.111 <sup>*</sup>	.155 <sup>**</sup>
3. Openness/Intellect	-.115	.155 <sup>*</sup>	-	.123	-.054
4. Agreeableness	-.206 <sup>**</sup>	.436 <sup>***</sup>	.190 <sup>*</sup>	-	.023
5. Conscientiousness	-.277 <sup>***</sup>	.303 <sup>***</sup>	-.108	.208 <sup>*</sup>	-

\* $p < .05$ ; \*\* $p < .01$ ; \*\*\* $p < .001$

Note: Estimates below the diagonal are from the CFA model. Estimates above the diagonal are from the ESEM model.

Table 4: Factor Inter-correlations for the BFI Model from CFA and ESEM Models

	1	2	3	4	5
1. Neuroticism	-	-.254***	.004	-.210**	-.269***
2. Extraversion	-.380***	-	.136*	.082	.119*
3. Openness/Intellect	-.222**	.244***	-	.066	-.008
4. Agreeableness	-.428***	.188**	.129	-	.205**
5. Conscientiousness	-.321***	..167**	-.033	.432***	-

\* $p < .05$ ; \*\* $p < .01$ ; \*\*\* $p < .001$

Note: Estimates below the diagonal are from the CFA model. Estimates above the diagonal are from the ESEM model.

Table 5: Correlations between CFA and ESEM Factor Scores and Factor Score Determinacy

Estimates

Inventory	Factor Labels				
	N	E	O	A	C
NEO-FFI					
<i>Correlation</i>	.984	.929	.982	.914	.986
<i>Shared Variance</i>	96.8%	86.3%	96.4%	83.5%	97.2%
Determinacies					
<i>CFA</i>	.943	.920	.900	.873	.939
<i>ESEM</i>	.948	.917	.902	.897	.948
BFI					
<i>Correlation</i>	.975	.992	.963	.922	.978
<i>Shared Variance</i>	95.1%	98.4%	92.7%	85.0%	95.7%
Determinacies					
<i>CFA</i>	.926	.942	.918	.882	.901
<i>ESEM</i>	.930	.951	.928	.884	.908

Table 6: NEO-FFI Parameter Estimates for the Regression of CFA and ESEM Factors on Criterion Variables

	TEI		Creativity	
	CFA	ESEM	CFA	ESEM
<i>Variance Explained</i>	<i>70.9%</i>	<i>71.2%</i>	<i>30.5%</i>	<i>29.5%</i>
Neuroticism	<b>-.511</b> (-.591 to -.430)	<b>-.571</b> (-.663 to -.479)	.064 (-.049 to .177)	-.009 (-.118 to .099)
Extraversion	<b>.277</b> (.186 to .368)	<b>.277</b> (.146 to .409)	<b>.388</b> (.250 to .526)	<b>.327</b> (.201 to .453)
Openness	<b>.164</b> (.084 to .243)	<b>.156</b> (.061 to .251)	<b>.338</b> (.238 to .438)	<b>.309</b> (.200 to .418)
Agreeableness	-.038 (-.124 to .048)	-.103 (-.323 to .117)	<b>.242</b> (.118 to .367)	.143 (-.057 to .343)
Conscientiousness	<b>.234</b> (.156 to .313)	<b>.288</b> (.196 to .381)	<b>.256</b> (.137 to .375)	<b>.236</b> (.110 to .361)

*Note:* Parameter estimates significant at the .05 level are shown in bold. All values are standardized. 95% confidence intervals are displayed in parentheses below point-estimates.

Table 7: BFI Parameter Estimates for the Regression of CFA and ESEM Factors on Criterion Variables

	Depression		Anxiety	
	CFA	ESEM	CFA	ESEM
<i>Variance Explained</i>	22.4%	22.7%	49.7%	49.8%
Neuroticism	<b>.400</b> (.282 to .518)	<b>.378</b> (.262 to .493)	<b>.771</b> (.677 to .864)	<b>.724</b> (.644 to .804)
Extraversion	<b>-.140</b> (-.259 to -.021)	<b>-.162</b> (-.275 to -.049)	.040 (-.060 to .141)	-.007 (-.110 to .096)
Openness	.021 (-.084 to .126)	-.031 (-.163 to .100)	.054 (-.044 to .152)	-.024 (-.172 to .124)
Agreeableness	-.094 (-.247 to .058)	-.062 (-.205 to .080)	-.073 (-.181 to .035)	-.029 (-.130 to .034)
Conscientiousness	-.102 (-.244 to .040)	-.101 (-.231 to .029)	.055 (-.047 to .157)	.067 (-.034 to .167)

*Note:* Parameter estimates significant at the .05 level are shown in bold. All values are standardized. 95% confidence intervals are displayed in parentheses below point-estimates.

Figures

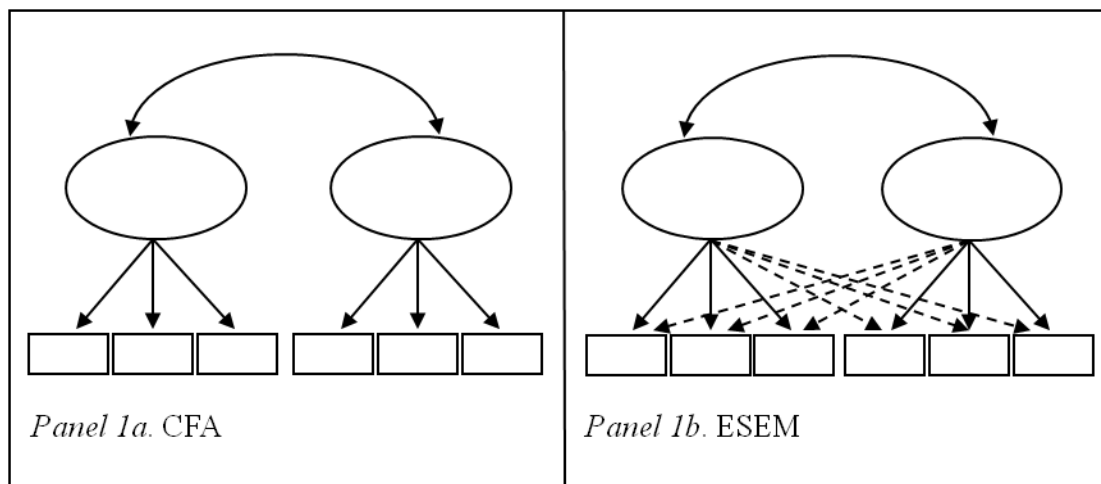


Figure 1. CFA and ESEM representations of a two-factor model

## **Online Supplementary Material**

Booth, T., & Hughes, D.J. (Submitted). Exploratory Structural Equation  
Modelling of Personality Data. *Assessment*.

The tables in this document present the comparable results to those reported in the main manuscript but based on weighted least squares means and variances adjusted estimation.



T1: Model Fit for Five Factor Models of the NEO-FFI (n=336) and BFI (n=425)

	$\chi^2$	<i>df</i>	Sig.	CFI	TLI	RMSEA	WRMR/ SRMR	AIC	BIC
<i>NEO-FFI</i>									
5 Factor CFA	3634.342	1700	<i>p</i> <.001	0.751	0.741	0.056	1.787	-	-
5 Factor ESEM	2311.687	1480	<i>p</i> <.001	0.893	0.872	0.039	0.953	-	-
<i>BFI</i>									
5 Factor CFA	2657.498	892	<i>p</i> <.001	0.824	0.813	0.068	1.852	-	-
5 Factor ESEM	1540.128	736	<i>p</i> <.001	0.920	0.897	0.051	0.939	-	-

T2: Factor Inter-correlations for the NEO-FFI Model from CFA and ESEM Models

	1	2	3	4	5
1. Neuroticism	-	-.188***	-.080	-.029	-.157**
2. Extraversion	-.480***	-	.100*	.110**	.163**
3. Openness/Intellect	-.182**	.167**	-	.110*	-.044
4. Agreeableness	-.290***	.493***	.210**	-	.031
5. Conscientiousness	-.276***	.300***	-.074	.274***	-

\* $p < .05$ ; \*\* $p < .01$ ; \*\*\* $p < .001$

Note: Estimates below the diagonal are from the CFA model. Estimates above the diagonal are from the ESEM model.

T3: Factor Inter-correlations for the BFI Model from CFA and ESEM Models

	1	2	3	4	5
1. Neuroticism	-	-.246***	.042	-.178**	-.292***
2. Extraversion	-.427***	-	.145**	.081	.143**
3. Openness/Intellect	-.204***	.285***	-	.055	.028
4. Agreeableness	-.462***	.229***	.173**	-	.207***
5. Conscientiousness	-.345***	.242***	-.006	.471***	-

\* $p < .05$ ; \*\* $p < .01$ ; \*\*\* $p < .001$

Note: Estimates below the diagonal are from the CFA model. Estimates above the diagonal are from the ESEM model.

T4: Correlations between CFA and ESEM Factor Scores

Inventory	Factor Labels				
	N	E	O	A	C
<i>NEO-FFI</i>					
<i>Correlation</i>	.975	.903	.981	.868	.985
<i>Shared Variance</i>	95.1%	81.5%	96.2%	75.3%	97.0%
<i>BFI</i>					
<i>Correlation</i>	.967	.984	.959	.884	.968
<i>Shared Variance</i>	93.5%	96.8%	92.0%	78.1%	93.7%

T5: NEO-FFI Parameter Estimates for the Regression of CFA and ESEM Factors on  
Criterion Variables

	TEI		Creativity	
	CFA	ESEM	CFA	ESEM
<i>Variance Explained</i>	69.8%	70.3%	29.6%	28.4%
Neuroticism	<b>-.495</b> (-.559 to -.431)	<b>-.559</b> (-.645 to -.473)	.071 (-.037 to .179)	-.013 (-.125 to .099)
Extraversion	<b>.255</b> (.179 to .331)	<b>.281</b> (.128 to .435)	<b>.397</b> (.266 to .529)	<b>.320</b> (.213 to .427)
Openness	<b>.176</b> (.112 to .241)	<b>.154</b> (.086 to .221)	<b>.321</b> (.222 to .419)	<b>.284</b> (.188 to .380)
Agreeableness	-.044 (-.118 to .030)	<b>-.137</b> (-.250 to -.023)	<b>.246</b> (.125 to .367)	.097 (-.026 to .220)
Conscientiousness	<b>.237</b> (.178 to .297)	<b>.301</b> (.232 to .370)	<b>.274</b> (.161 to .387)	<b>.254</b> (.143 to .366)

*Note:* Parameter estimates significant at the .05 level are shown in bold. All values are standardized. 95% confidence intervals are displayed in parentheses below point-estimates.

T6: BFI Parameter Estimates for the Regression of CFA and ESEM Factors on Criterion

Variables

	Depression		Anxiety	
	CFA	ESEM	CFA	ESEM
<i>Variance Explained</i>	20.9%	21.4%	46.2%	45.7%
Neuroticism	<b>.391</b> (.286 to .496)	<b>.364</b> (.267 to .462)	<b>.755</b> (.672 to .837)	<b>.690</b> (.621 to .758)
Extraversion	<b>-.142</b> (-.257 to -.027)	<b>-.181</b> (-.280 to -.082)	.042 (-.050 to .134)	-.036 (-.124 to .052)
Openness	.045 (-.053 to 0.143)	-.019 (-.122 to .084)	.051 (-.032 to .133)	-.040 (-.148 to .068)
Agreeableness	-.086 (-.211 to .039)	.057 (-.033 to .148)	-.088 (-.184 to .009)	0.21 (-.055 to .097)
Conscientiousness	-.076 (-.213 to .060)	-.070 (-.176 to .037)	.056 (-.043 to .155)	.083 (-.029 to .168)

*Note:* Parameter estimates significant at the .05 level are shown in bold. All values are standardized. 95% confidence intervals are displayed in parentheses below point-estimates.